caBIG – the end of the beginning…

Ken Buetow
NCICB/NCI/NIH/DHHS

NATIONAL CANCER INSTITUTE

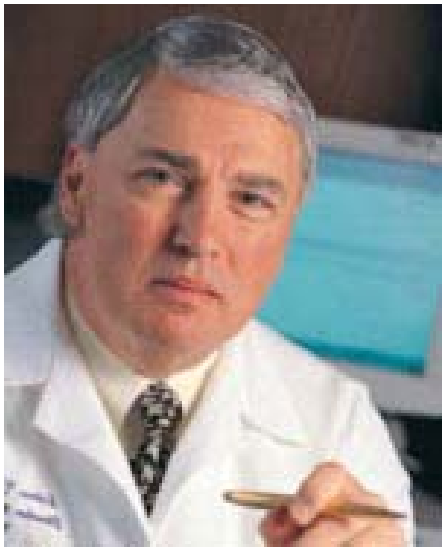
The NCI 2015 challenge goal:

… eliminate death and suffering due to cancer

"When I look into the eyes of a patient losing the battle with cancer, I say to myself, It doesn't have to be this way."

Dr. A.C. von Eschenbach, M.D.
Director, National Cancer Institute

The Nation's Investment in Cancer Research (2003)



NATIONAL CANCER INSTITUTE

NCI biomedical informatics

- Goal: A virtual web of interconnected data, individuals, and organizations redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise

NATIONAL CANCER INSTITUTE

Cancer Biomedical Informatics Grid (caBIG)

- Common, widely distributed infrastructure permits cancer research community to focus on innovation
- Shared vocabulary, data elements, data models facilitate information exchange
- Collection of interoperable applications developed to common standard
- Raw published cancer research data is available for mining and integration

NATIONAL CANCER INSTITUTE

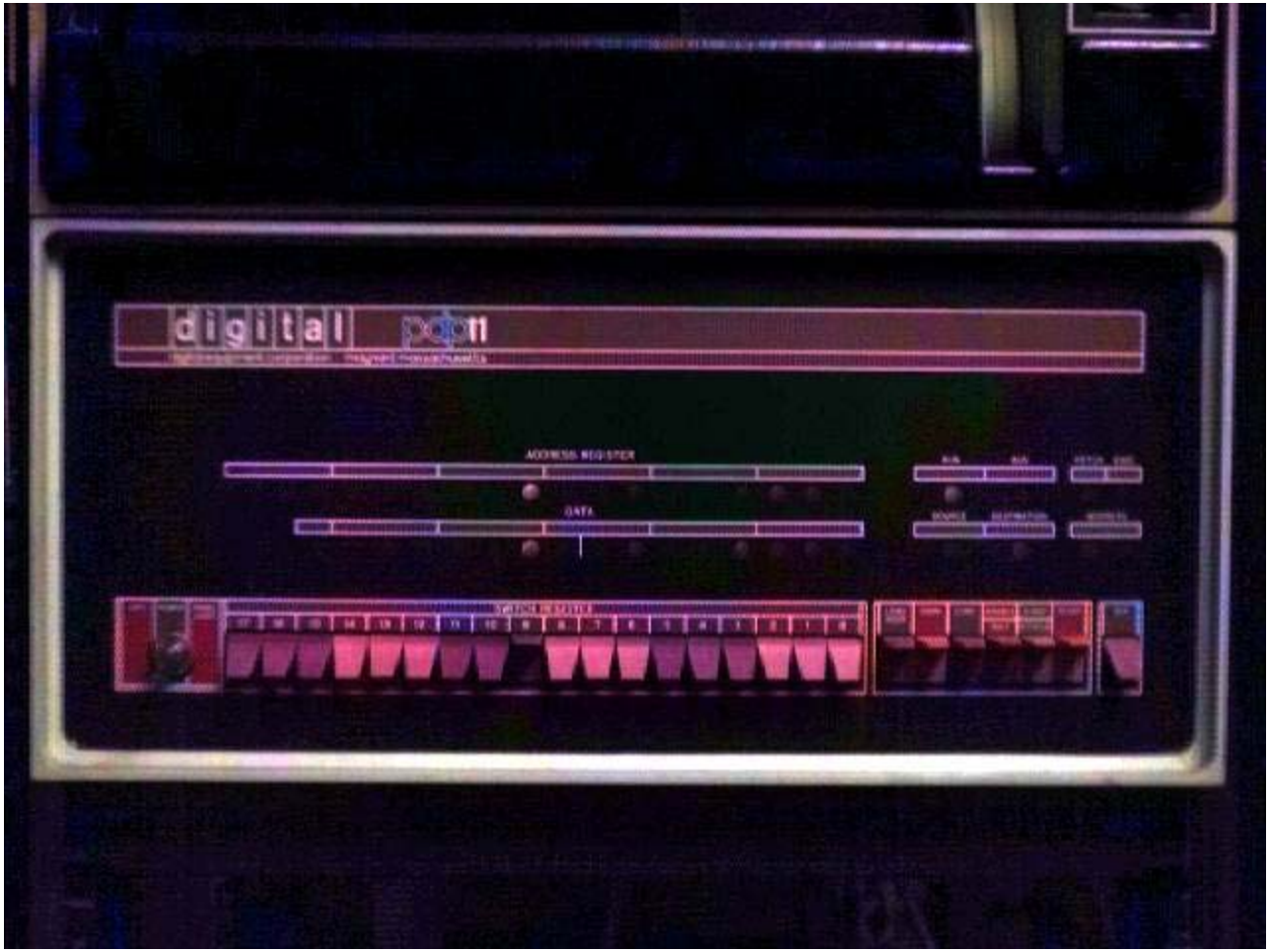
caBIG principles

- Open source
- Open access
- Open development
- Federated

Goals of the caBIG pilot:

- Illustrate that a *spectrum of Cancer Centers with varying needs and capabilities can be joined* in a common network of communications, shared data, applications, and technologies
- Demonstrate that Cancer Centers, in collaboration with NCI, *will develop new enabling tools and systems* that could support multiple Cancer Centers
- Demonstrate that Cancer Centers will actively *use the grid and realize greater value* in their cancer research endeavors by using the network
- Create an *extensible infrastructure* that will continue to be *expanded and extended to members of the cancer research community*

caBIG "power-up"



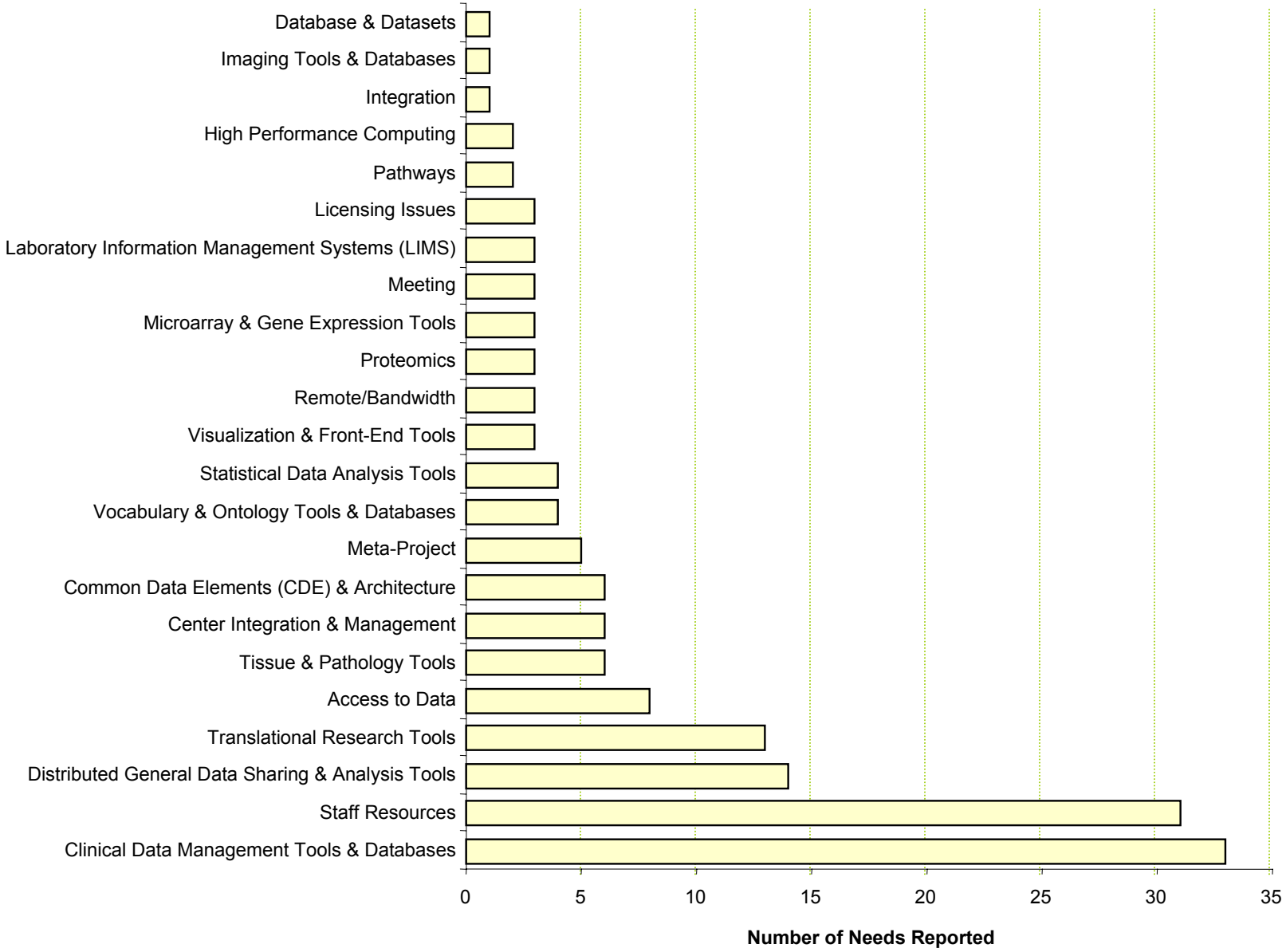
initiating the caBIG ***BIOS***

- A power-on self-test (POST) for all of the different hardware components in the system to make sure everything is working properly
- Activating other BIOS chips on different cards installed in the computer
- Providing a set of low-level routines that the operating system uses to interface to different hardware devices
- Managing a collection of settings for the hard disks, clock, etc.

(http://computer.howstuffworks.com/bios)

caBIG community interaction

- Informational seminars
  - over 100 participants
- Cooperative Development Meetings
  - visited 50 Cancer Centers
- caBIG project summaries
  - 118 pilot projects nominated
- caBIG WWW site
  - On-line forum: Message boards/file sharing
- caBIG Program Updates


Database & Datasets
Imaging Tools & Databases
Integration
High Performance Computing
Pathways
Licensing Issues
Laboratory Information Management Systems (LIMS)
Meeting
Microarray & Gene Expression Tools
Proteomics
Remote/Bandwidth
Visualization & Front-End Tools
Statistical Data Analysis Tools
Vocabulary & Ontology Tools & Databases
Meta-Project
Common Data Elements (CDE) & Architecture
Center Integration & Management
Tissue & Pathology Tools
Access to Data
Translational Research Tools
Distributed General Data Sharing & Analysis Tools
Staff Resources
Clinical Data Management Tools & Databases
0
5
10
15
20
25
30
35
Number of Needs Reported

caBIG – *Operating System (v1.0)*

- **Processor management** - Breaking the tasks down into manageable chunks and prioritizing them before sending to the CPU
- **Memory management** - Coordinating the flow of data in and out of RAM and determining when virtual memory is necessary
- **Device management** - Providing an interface between each device connected to the computer, the CPU and applications
- **Storage management** - Directing where data will be stored permanently on hard drives and other forms of storage
- **Application Interface** - Providing a standard communications and data exchange between software programs and the computer
- **User Interface** - Providing a way for you to communicate and interact with the computer

(http://computer.howstuffworks.com)

The caBIG pilot structure has been developed to capitalize on existing Cancer Center enthusiasm by maximizing Cancer Center involvement while addressing areas of critical need

caBIG pilot management structure is centered around Workspaces, Working Groups, and an associated Management Structure for these groups

- **Workspaces** – Cancer Centers will be grouped together in areas called "Workspaces." Each Workspace will encompass projects or activities with a common focus

- **Working Groups** – these groups will provide support to the caBIG pilot Workspaces and project teams. Strategic level planning groups will also be convened to support the overall caBIG initiative

- **Management Structure** – the structure will incorporate three levels of management; Governance level, Intermediate level, and Workspace/ Project level management

NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

Three Domain Workspaces and two Cross Cutting Workspaces are being launched during boot cycle

DOMAIN WORKSPACE 1
Clinical Trial Management Systems

Will address the need for consistent, open components for clinical trials management. Will capitalize on existing tools and capabilities.

DOMAIN WORKSPACE 2
Integrative Cancer Research

Will provide tools to enable collection, processing, sharing and integration of cancer research data.

DOMAIN WORKSPACE 3
Tissue Banks & Pathology Tools

Will provide for the integration, development, and implementation of tissue and pathology tools. Will leverage current systems and related resources.

Will be responsible for evaluating, developing, and integrating systems for vocabulary and ontology content, standards, and software systems for content delivery. Will establish and emphasize common standards.

CROSS CUTTING WORKSPACE 1
Vocabularies & Common Data Elements

Will develop architectural standards and provide architectural assistance as necessary to other Workspaces. Will provide a common framework and draw on the lessons learned from related novel initiatives.

CROSS CUTTING WORKSPACE 2
Architecture

NATIONAL CANCER INSTITUTE

Three Strategic Level Planning Groups are being launched during boot cycle

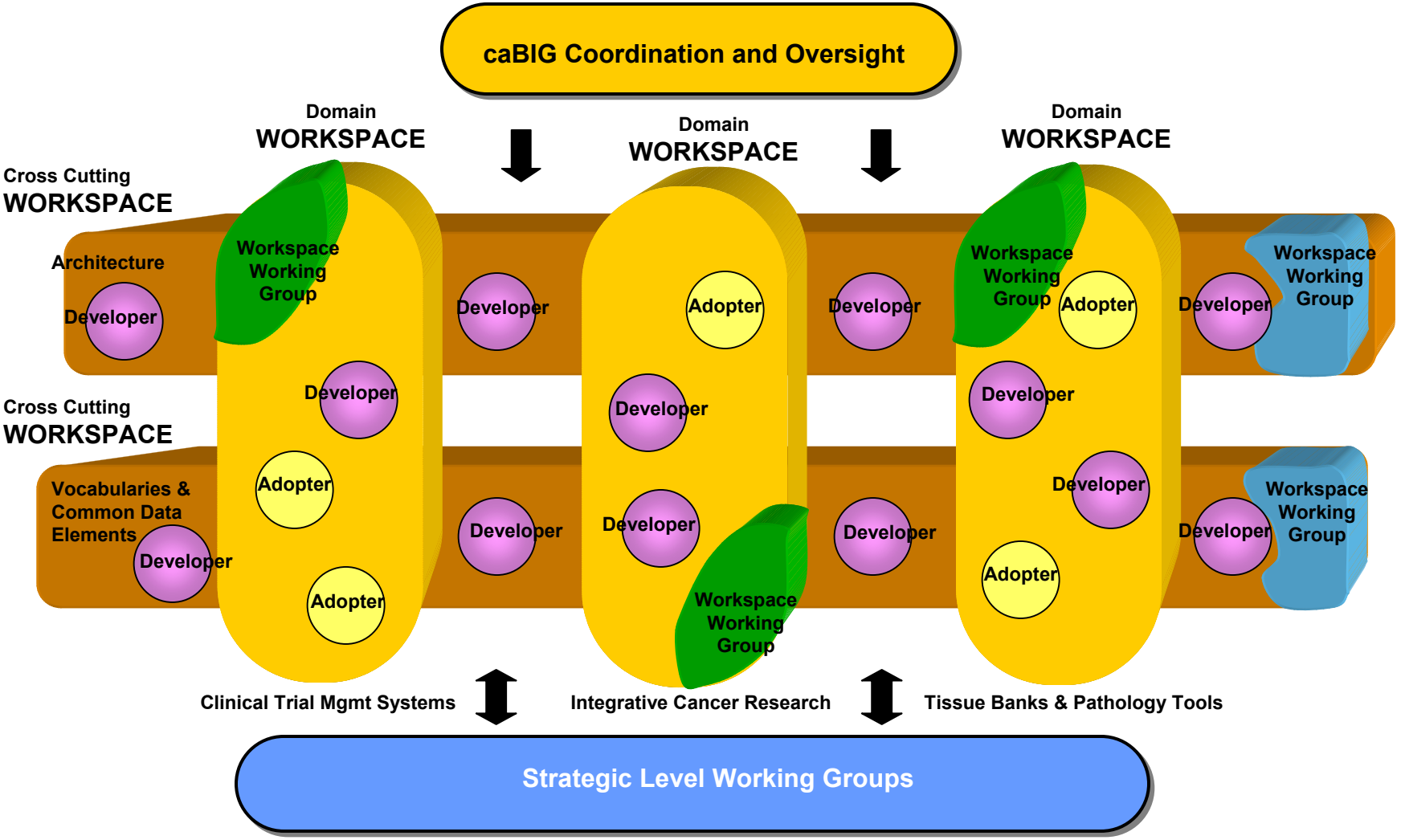
Data Sharing and Intellectual Capital	Addresses issues related to the sharing of data, applications and infrastructure both within the consortium and in the larger cancer research community.
Training	Develops strategies for providing training in the use of the caBIG developed resources including on-line tutorials, workshops, training programs.
caBIG Strategic Planning	Assists in identifying strategic priorities for the development and evolution of the caBIG effort.

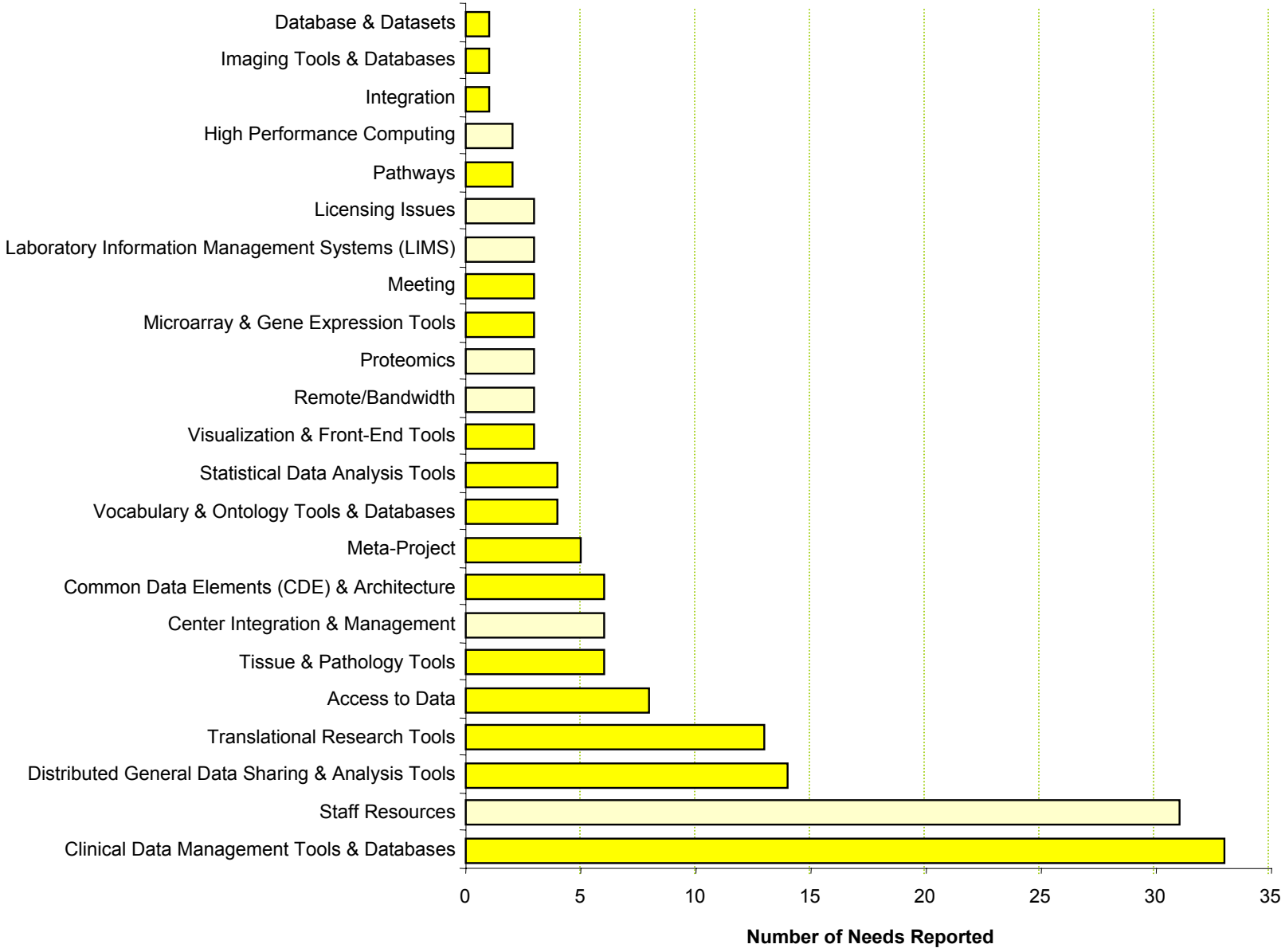
Industry Grade Software Engineering Processes

- Software engineering activities will be conducted according to CMM based industry best practices.
- Use of iterative software development methodology to support component based engineering.
- Will encompass proactive project planning, implementation, monitoring, testing, adoption, and documentation
    - Metrics driven project management approach
    - Project plans with defined deliverables
    - Performance based monitoring
- Workspaces and Working Groups participate in developing processes and artifacts.
- The Master Contractor (NCI appointed contractor) will coordinate and monitor Workspace activities according to these defined processes.

caBIG cancer Biomedical Informatics Grid

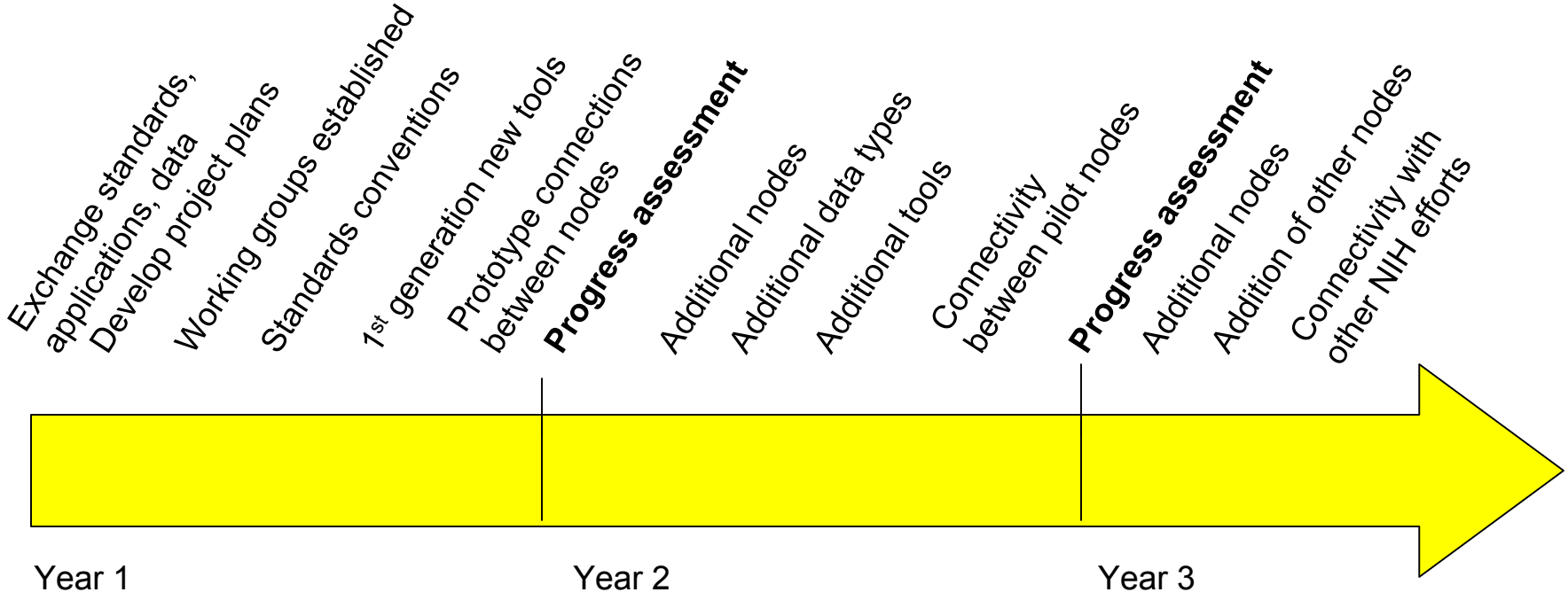
caBIG – *OS v1.0*




Database & Datasets
Imaging Tools & Databases
Integration
High Performance Computing
Pathways
Licensing Issues
Laboratory Information Management Systems (LIMS)
Meeting
Microarray & Gene Expression Tools
Proteomics
Remote/Bandwidth
Visualization & Front-End Tools
Statistical Data Analysis Tools
Vocabulary & Ontology Tools & Databases
Meta-Project
Common Data Elements (CDE) & Architecture
Center Integration & Management
Tissue & Pathology Tools
Access to Data
Translational Research Tools
Distributed General Data Sharing & Analysis Tools
Staff Resources
Clinical Data Management Tools & Databases
0
5
10
15
20
25
30
35
Number of Needs Reported

caBIG cancer Biomedical Informatics Grid

caBIG milestones/deliverables



NATIONAL CANCER INSTITUTE

caBIG Kick-off goals

- Finish the boot cycle
  - Complete and share inventory of caBIG components
  - Determine requirements, scope, sequence for components
- Put in place workspace mechanics for modifying/constructing additional caBIG components
- Match caBIG infrastructure/applications to early adopters
  - Adopters determine testing/data to be made available to caBIG community
- Establish working group processes for moving forward

caBIG - OS v1.0 overview

- Users guide - Mark Adams
- Technical Guide – Peter Covitz
  - Current specifications
  - Extending the specification
- Applications Developers Guide – Sue Dubman
  - Pre-loaded applications
  - Generating new applications

NATIONAL CANCER INSTITUTE